Preprint
UCRL-JC-143945

LLNL Middle East, North Africa and Western Eurasia Knowledge Base

J.L. O'Boyle, S.D. Ruppert, T.F. Hauk, D.A. Dodge, F. Ryall, M.A. Firpo

This article was submitted to
23rd Annual Department of Defense/Department of Energy Seismic Research Review, Jackson Hole, WY, October 1-5, 2001

July 12, 2001

LLNL MIDDLE EAST, NORTH AFRICA AND WESTERN EURASIA KNOWLEDGE BASE

Jennifer L. O'Boyle, Stanley D. Ruppert, Teresa F. Hauk, Douglas A. Dodge,
Flori Ryall, and Michael A. Firpo and LLNL GNEM Staff

Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. W-7405-ENG-48

ABSTRACT

The Lawrence Livermore National Laboratory (LLNL) Ground-Based Nuclear Event Monitoring (GNEM) program has made significant progress populating a comprehensive Seismic Research Knowledge Base (SRKB) and deriving calibration parameters for the Middle East, North Africa and Western Eurasia (ME/NA/WE) regions. The LLNL SRKB provides not only a coherent framework in which to store and organize very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data, but also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces. The SRKB is a flexible and extensible framework consisting of a relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools. This SRKB framework is designed to accommodate large volumes of data (almost 3 million waveforms from 57,000 events) in diverse formats from many sources (both LLNL derived research and integrated contractor products), in addition to maintaining detailed quality control and metadata. We have developed expanded look-up tables for critical station parameter information (including location and response) and an integrated and reconciled event catalog data set (including specification of preferred origin solutions and associated phase arrivals) for the PDE, CMT, ISC, REB and selected regional catalogs. Using the SRKB framework, we are combining travel-time observations, event characterization studies, and regional tectonic models to assemble a library of ground truth information and phenomenology (e.g. travel-time and amplitude) correction surfaces required for support of the ME/NA/WE regionalization program. We also use the SRKB to integrate data and research products from a variety of sources, such as contractors and universities, to merge and maintain quality control of the data sets. Corrections and parameters distilled from the LLNL SRKB provide essential contributions to the National Nuclear Security Administration Knowledge Base (NNSA KB) for the ME/NA/WE region and will improve capabilities for underground nuclear test monitoring. The LLNL research products will facilitate calibration of International Monitoring System (IMS) stations (primary and auxiliary), their surrogates (if not yet installed) and selected gamma stations necessary to complete the ME/NA/WE regionalization efforts. In addition to an overview of selected individual research products, we present an overview of our visualization, integration, validation, and organizational processes. Development of these processes and the LLNL SRKB was necessitated by both the very large amount of data and information involved, over 2 terabytes, and the varied data and research result formats utilized. Products assembled, integrated and validated using the LLNL SRKB are grouped into 5 major categories:

1. Reference and contextual information
2. Detection data
3. Calibration and ground truth data
4. Event location products
5. Event identification products

KEY WORDS: seismic, database, knowledge base, GIS, visualization, integration

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

OBJECTIVE

The primary objective of the Lawrence Livermore National Laboratory (LLNL) Seismic Research Knowledge Base (SRKB) is to integrate the LLNL Ground-Based Nuclear Explosion Monitoring (GNEM) Middle East, North Africa and Western Eurasia (ME/NA/WE) regionalization program. The LLNL SRKB provides efficient access to, and organization of, thousands of seismic events and associated waveforms, while also providing the framework to store, organize, integrate and disseminate research results for delivery into the National Nuclear Security Administration Knowledge Base (NNSA KB).

The SRKB provides a unified framework for all seismic data and research products as outlined in Figure 1. This requires the reconciliation and merging of data derived from different sources and methods and of varying quality, along with the ranking of the data sets based on relative quality. Integration occurs on many levels in order to generate data, data sets, databases or knowledge bases. The SRKB also allows for the integration of LLNL research with contractor research products and other available data sets. In order to efficiently organize information within the SRKB, it was necessary to automate procedures needed to create and update database tables, but a large effort is still required by technicians and scientists to load special data sets, review results of automated processing and resolve quality issues. Sufficient metadata (including measurement procedures, codes, comments and measurement errors) are stored at each step in the data creation and analysis process to allow re-creation or verification of results at any stage in the processing flow.

LLNL research products created using the SRKB may be grouped under two major categories: primary data products and derived products. The primary products are those developed in the process of collecting the raw materials for calibration: ground truth data, waveform data, event catalogs, phase pick information, regional station information and instrument responses. The derived products (distilled from the organized raw seismological data) are models and corrections that improve detection, location and identification functions. To calibrate International Monitoring System (IMS) seismic stations (primary and auxiliary), as well as a variety of gamma stations that experience has shown to be useful, the LLNL SRKB must incorporate and organize the following categories of primary and derived measurements, data and metadata:

Contextual and Raw Data
1. Station Parameters and Instrument Responses
2. Global and Regional Earthquake Catalogs
3. Selected Calibration Events
4. Event Waveform Data
5. Geologic/Geophysical Data sets
6. Geophysical Background Model

Measurements and Research Results
1. Phase Picks
2. Travel-time and Velocity Models
3. Rayleigh and Love Surface Wave Group Velocity Measurements
4. Phase Amplitude Measurements and Magnitude Calibrations
5. Detection and Discrimination Parameters

Corrections and parameters distilled from the LLNL SRKB provide needed contributions to the NNSA KB for the ME/NA/WE region and will enable the United States National Data Center (US NDC) to effectively monitor for nuclear explosions. The LLNL contributions support critical US NDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review. Figure 2 outlines the processes of data collection, research and integration that result in contributions to the NNSA KB.

RESEARCH ACCOMPLISHED

The LLNL SRKB is a framework consisting of an ORACLE relational database (RDB), Geographical Information System (GIS), and associated product/data visualization and data management tools. The SRKB is necessary for the storage and organization of very large volumes of collected seismic waveforms, associated event parameter information, and spatial contextual data, but also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces. Figure 3 outlines the interaction between all components of the SRKB. The relational database is organized in CSS3.0 (Center for Seismic

Studies Version 3.0) format with US NDC and LLNL extensions. This format provides parameter defined tables for different elements of seismic data, such as event and station information, as well as allowing for customized tables to be developed for specific research needs or results and the compatibility with other organizations provides for easy delivery and integration of products.

Contextual and Raw Data

Station Parameters

Seismic station information is a metadata requirement needed to support all stages of seismic waveform analysis. These metadata include such parameters as station operation dates, location and elevation, type of channels and instruments, sampling rates, and instrument responses. Our main source of this information is Incorporated Research Institutes for Seismology (IRIS) "dataless" SEED files, which are provided by each of the networks affiliated with IRIS. Other station information has been obtained through Internet station books and AutoDRM systems. CSS3.0 **site** and **sitechan** table entries (listing station location, available channels, sensor orientations, operation dates, etc.) were created for almost all IRIS affiliated networks as well as many other stations with waveforms in the SRKB. All station parameters are reviewed among existing information sources, and conflicts must be resolved between sources and in reference to waveforms available before database tables are updated.

Network codes are appended to station codes, and location codes are appended to channel names when multiple stations and/or instruments are concurrently operating with the same name. Over 1800 station and array element table entries have been updated, but reliable parameter data do not exist for some stations if only minimal or inconsistent information is available. **Instrument** and **sensor** tables are used to document instrument type and response for each station and channel. The Seismic Analysis Code (SAC) has been modified to interpret response information in RESP or FAP format for use in performing instrument response corrections on waveform data by using the EVALRESP software library available from IRIS.

Event Bulletins

Reference event locations and origin time information are necessary in most stages of our seismic processing and research. Bulletin information from many global, local, and regional earthquake catalogs has been incorporated into the LLNL SRKB and provides a much larger source for event selection. This combined and reconciled source facilitates comparison of event parameters provided by multiple networks with different degrees of location accuracy and provides a wider range of magnitudes and event types. The global catalogs include: United States Geological Survey (USGS) Monthly (Final) Preliminary Determination of Epicenters (PDE) catalog, Bulletin of the International Seismological Centre (ISC), Harvard Centroid Moment Tensor (CMT) catalog, and the International Data Centre Reviewed Event Bulletin (REB). We have also compiled numerous regional and local catalogs for the ME/NA/WE region, including Jordan, Israel, Scandinavia and the Kola Peninsula. A special catalog was compiled for events in the European Arctic and Novaya Zemlya in collaboration with NORSAR. Figure 5 shows a map of the event location coverage provided by several of the bulletins in the ME/NA/WE region. We have established a number of collaborative agreements with countries and institutes in our study region to obtain both local seismic catalogs and ground truth information as well as seismic waveform data.

Waveforms

We are collecting seismic data from IMS primary and auxiliary stations, as well as surrogate stations (for IMS stations not yet installed) and other stations needed to support calibration in the region of study. We have obtained up to 10 years of continuous data for important ME/NA/WE stations from IRIS, Institut de Physique du Globe de Paris GEOSCOPE program, the GEO-Forschungs Zentrum/Potsdam, Germany GEOFON program and other data centers. Current data are being supplied by an LLNL joint project with the Jordan Natural Resources Authority from two seismometers deployed in Jordan. Data for particular events has been obtained from the prototype International Data Center (PIDC) and the US NDC. The Center for Monitoring Research (CMR) and NORSAR have provided waveforms for special regions/events, such as the Novaya Zemlya test site. Stations in the ME/NA/WE region with event waveforms stored in the SRKB are shown in Figure 4.

Most of the archived waveforms in the LLNL SRKB are from events located within the ME/NA/WE region and occurring between 1976 and 2001. Our current data collection emphasis is on events recorded between 1990 and the present, except for IMS surrogate stations no longer in operation or for special events such as nuclear tests. Although the continuous data remain archived on tapes, data segments for seismic events are extracted

from the continuous waveforms. The number of waveforms in the SRKB is now almost 3 million, which represents 57,000 seismic events. In addition to individual event waveform segments and continuous data traces managed by the SRKB, we also maintain an archive of active and passive seismic data from various field deployments, which include IRIS/PASSCAL experiments in Tanzania, Pakistan, Caspian Sea, and Turkmenistan. We are also archiving the very long Former Soviet Union Peaceful Nuclear Explosion (PNE) refraction profiles collected under contract to LLNL by the USGS.

Other Reference and Ground Truth Data sets

Projects to develop reference data sets include waveform correlation and subspace detectors to provide statistics on mining activity (see the LLNL Detection Program abstract (Harris *et al*., this Proceedings) and a regionalization model based on *a priori* geophysical knowledge used as a reference model for model-based correction surfaces. Collaborations with Cornell University, USGS and other organizations allow us to obtain reference data sets useful as background and supporting information for research. Most of these data sets are provided as Geographic Information System (GIS) products, which allow us to integrate them with LLNL data sets. The data sets include geographic, geopolitical, mining industry, geologic and tectonic information for the ME/NA/WE region. We also collect photographs and satellite imagery to support ground truth determination.

Measurements and Research Results

Phase Information

LLNL researchers and analysts have made phase picks for over 1600 events to yield over 20,000 travel-time observations available to the LLNL research team for location and discrimination projects. Phase information is recorded in the **arrival** and **assoc** tables, along with pick and waveform quality and other comments that are recorded in a **remark** table. Augmenting the LLNL picks, we added 30 million ISC phase arrival measurements from 1964-1998 and almost 2 million REB arrivals to be used for travel-time correction studies. Phase picks have also been entered into the SRKB from a reviewed subset of the ISC bulletin (Engdahl *et al*., 1998) and a few Scandinavian bulletins.

Seismic Location Research

Seismic location researchers utilize event catalogs (especially ground truth), waveforms and phase picks to develop travel-time corrections. Improvement in seismic location is accomplished by combining model-based and empirical travel-time corrections. After a calibration model is applied, empirical corrections are calculated using Modified Bayesian Kriging algorithm with travel-time residuals for suitably well located calibration events. The travel-time and velocity models developed through this research are an important part of the SRKB. See the LLNL Calibration Program (Myers *et al*., this Proceedings) and LLNL Seismic Location (Schultz *et al*., this Proceedings) papers for more details about these aspects of the LLNL GNEM program research.

Event Identification and Magnitude Research

The event identification researchers also use event catalogs, waveforms and phase picks to develop products that provide seismological information to improve the ability to identify underground nuclear explosions, by discriminating them from the background of earthquakes. The magnitude research results in calibration products and magnitudes for 1-Hz regional phases, broadband coda waves, and long-period surface waves. This provides the necessary calibration information for regional data in order to calculate seismic magnitudes for use by nuclear monitoring functions (detection, location, identification, and characterization). Custom tables in the SRKB database store measurements and related parameters for surface wave group velocities, body wave amplitudes and magnitudes. See the LLNL Identification Program (Walter *et al*., this Proceedings) and LLNL Calibration Program (Myers *et al*., this Proceedings) papers for more details about these aspects of the LLNL GNEM program research.

Validation and Testing

The SRKB provides the framework to perform end-to-end process validation and testing of research results. Validation and assessment tasks include the definition and evaluation of specific metrics, analysis and use of metrics to define the stopping criteria for particular location and identification tasks, and ranking and performance evaluation of calibration activities. Success and efficient implementation of validation and assessment tasks require a framework that ties research results to data and measurements used to create each calibration product and that also supports statistical and visualization tools for performing the validation tasks. The metadata inherent in the SRKB allow multiple realizations of calibration products, created with different

processes and assumptions, to be created, stored, and compared with one another. Thus the SRKB facilitates the distillation, documentation, and delivery of research products.

Data Utilities

Data Input

We have written a set of C++ classes designed primarily as components of computer programs for use by the LLNL GNEM program, which consists of about 130 classes. Of these, about 10 are service classes with no specific seismological capability, *e.g.* time class, architecture class, etc. All the remaining classes are specialized to some degree for use in seismological applications. All of the collection classes have database-aware specializations as well, since most of the programs developed with the classes interact extensively with the SRKB database. The class library has built-in support for the CSS3.0, SAC, GSE2.0, and PC-SUDS data formats. Using the classes has proven to be an effective means of building the utility programs required to support the SRKB as described in the following paragraphs.

To manage and update the large number of response files for stations maintained in the LLNL SRKB, we have automated the required processing and quality control steps. As we load information into the **instrument** and **sensor** tables, the automated system separates the RESP file into individual epochs, removes overlapping epochs and verifies that the epochs correspond to those in the station parameter tables. New RESP files are created from these results with the station and channel names modified if necessary.

We have developed an automated method to load various bulletins of different formats into native-format database tables, which retain many of the fields provided in the original bulletin, as well as assign a unique origin identifier (*orid*) to each event. Individual bulletins are then merged into a single CSS3.0 **origin** table retaining the author of the original bulletin. A space-time correlation algorithm is used to reconcile events between individual catalogs. Events common to multiple bulletins will be assigned common event identifiers (*evid*) while retaining unique *orid*s. The **event** table stores the preferred origin for each event based on a ranked list of catalog preferences. Phase arrivals provided with native catalogs are loaded into the **assoc** and **arrival** tables.

To generate the event waveforms for the SRKB, we have developed an automated way of extracting waveforms based on events selected from an **origin** table and adding entries for these waveforms into the appropriate database tables. Waveforms are extracted from continuous data in the native format and then converted to CSS3.0 format. Currently, we can perform waveform extraction on data in SAC, SEED, CSS, SEGY and GSE formats. Information already in the database is used to determine station parameters and event origins, which are then used in determining the appropriate time window of data to extract. The program checks waveforms for poor or missing signals, which are then recorded in a metadata table. If waveforms for a specific event are segmented, the segments will be merged together; if a gap exists between segments, the space will be filled with zero-value data points. If two segments overlap, a correlation test is done before the two segments are joined.

Data Access

Different researcher needs for data and metadata require that subsets of data be provided in a format easily accessible to many diverse types of software and analysis tools. Therefore, the SRKB access tools have been designed to utilize the power of the relational database to facilitate efficient queries and data retrieval. The Seismic Analysis Code (SAC) software used by LLNL researchers provides direct access to database table information and waveforms and uses the response files to perform instrument response corrections. PL/SQL language can be used to make database queries on contents of any of the available database tables.

For spatial queries and organization, we have adopted a Geographic Information System (GIS), which provides a framework to store and manipulate spatially defined data. The GIS is linked with the ORACLE database to provide joint spatial and relational queries. We have integrated a large number of our research products and contextual data sets into the GIS. A customized framework provides an organized menu system for accessing the data sets and also includes additional functionality developed in collaboration with Sandia National Laboratories. Researchers can use the GIS to browse many of their research products and also to perform basic comparisons, queries, and analyses. It also provides a way to use LLNL products in conjunction with other data sets, including integrated contractor products.

Given the large quantity of data now managed by the SRKB, the need arose to create the efficient "production" level seismic data selection, processing and visualization tools necessary to meet programmatic and NNSA KB needs. These tools, along with data browsers to allow visualization and quick access to both data and delivered research products, were developed in collaboration with Sandia National Laboratories. The tools provide such functionality as deriving location corrections, making amplitude and magnitude measurements, and developing discriminants.

CONCLUSIONS AND RECOMMENDATIONS

The data sets and research products contained and organized within the LLNL SRKB may be grouped under 5 major categories:

1. Reference and Contextual Information
2. Detection Data
3. Calibration and Ground Truth Data
4. Event Location Products
5. Event Identification Products

The SRKB allows for the collection of raw and contextual seismic data to be used in research, provides an interface for the researchers to access data, provides a framework to store research results and integrate contractor data sets, and supports assembly and dissemination of data sets to the NNSA KB. A wide range of research products required to support the ME/NA/WE regionalization program are being derived from waveforms, station parameters, and event origin information contained in the LLNL SRKB. Corrections and parameters derived, assembled, integrated and validated using the LLNL SRKB provide essential contributions to the NNSA Knowledge Base for the ME/NA/WE region and will enable the US NDC to effectively monitor for nuclear explosions. The LLNL portion of the NNSA KB supports critical US NDC pipeline functions in detection, location, feature extraction, identification, and analyst review in the Middle East, North Africa and Western Eurasia.

We have made a major effort to provide a product development environment that encourages the natural synergies among each of the separate research efforts of the LLNL GNEM researchers. The derivation of reference and ground truth data sets and location and identification products takes place in an integrated environment with changes and improvements in one area being used to facilitate development of the remaining areas. We utilize the LLNL SRKB as an integrating framework to provide the basis for synergistic development of all LLNL GNEM research. By combining travel-time observations, event characterization studies, and regional wave-propagation studies of the LLNL GNEM research team for ground truth and regional events, we have assembled a substantial library of ground truth information (origin times, locations, depths, magnitudes), mine explosion statistics, tomographic models and travel-time and body-wave correction surfaces.

REFERENCES

Engdahl, R., R. van der Hilst, and R. Buland (1998), Global teleseismic earthquake relocation with improved travel-times and procedures for depth determination, *Bull. Seismol. Soc. Am.*, **88**, 722-743.

Harris, D., W. Walter, A. Rodgers, A. Sicherman, S. Myers, and C. Schultz (2001), LLNL Detection Program: Broad Area Characterization of Phase Detectability and Empirical Detectors for Specific Sources, 23rd Annual DTRA/NNSA Seismic Research Review Proceedings.

Myers, S., K. Mayeda, W. Walter, C. Schultz, A. Rodgers, A. Hofstetter, J. O'Boyle, and S. Ruppert (2001), LLNL Calibration Program: Data Collection, Ground Truth Validation, and Regional Coda Magnitude, 23rd Annual DTRA/NNSA Seismic Research Review Proceedings.

Schultz, C., M. Flanagan, F. Ryall, S. Myers, W. Hanley, J. Swenson, D. Dodge and M. Pasyanos (2001), LLNL Seismic Location: Validating Improvement Through Integration of Regionalized Models and Empirical Corrections, 23rd Annual DTRA/NNSA Seismic Research Review Proceedings.

Walter, W., A. Rodgers, M. Pasyanos, K. Mayeda, A. Sicherman, and D. Harris (2001), LLNL Identification Program: Regional Body-Wave Correction Surfaces and Surface-Wave Tomography Models to Improve Discrimination, 23rd Annual DTRA/NNSA Seismic Research Review Proceedings.

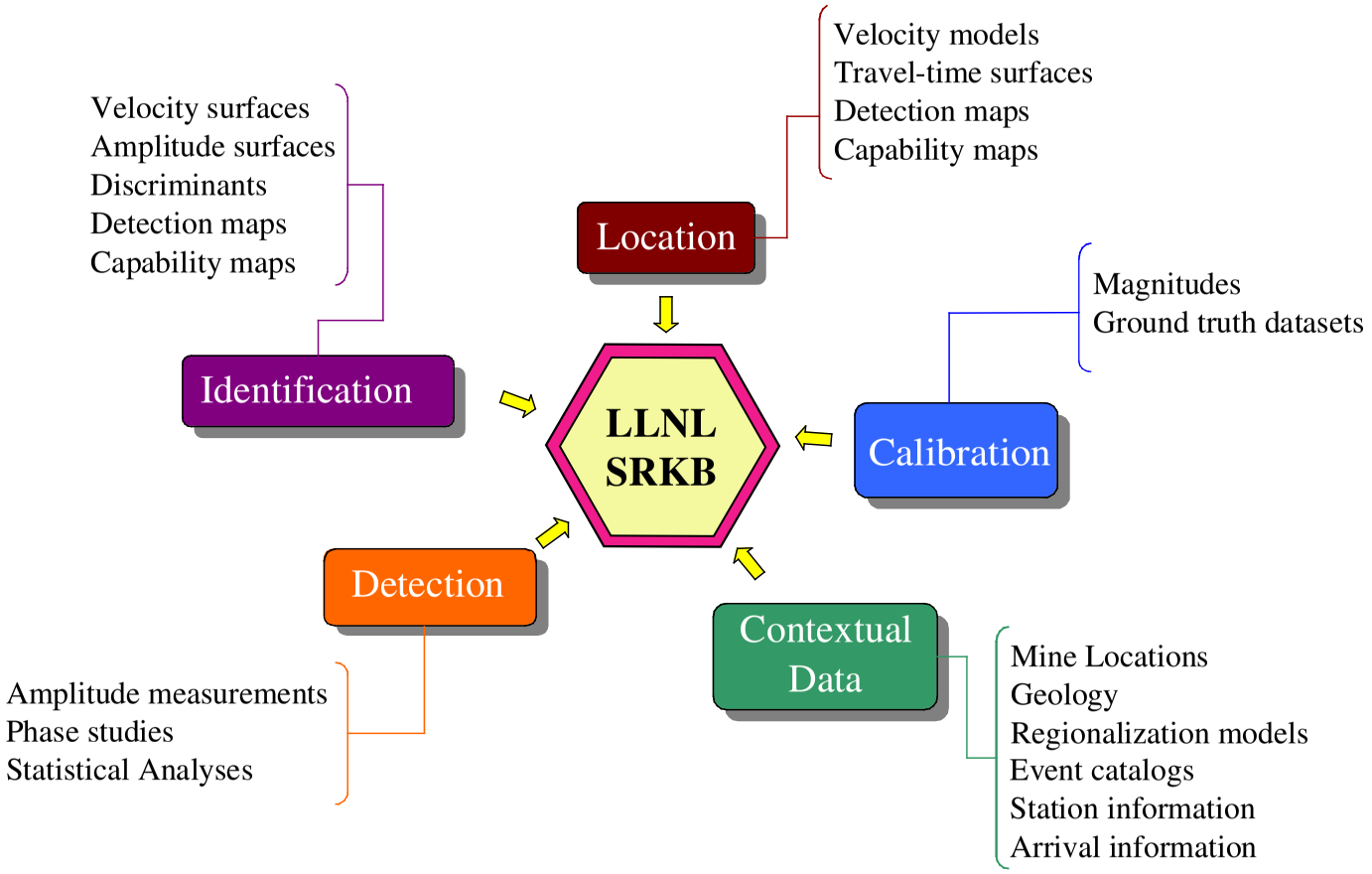

Figure 1. The LLNL SRKB provides a unified framework for all contextual/reference data and research products. The SRKB provides efficient access to, and organization of, thousands of seismic events and associated waveforms, while also providing the framework to store, organize, integrate and disseminate research results for delivery into the National Nuclear Security Administration Knowledge Base (NNSA KB).

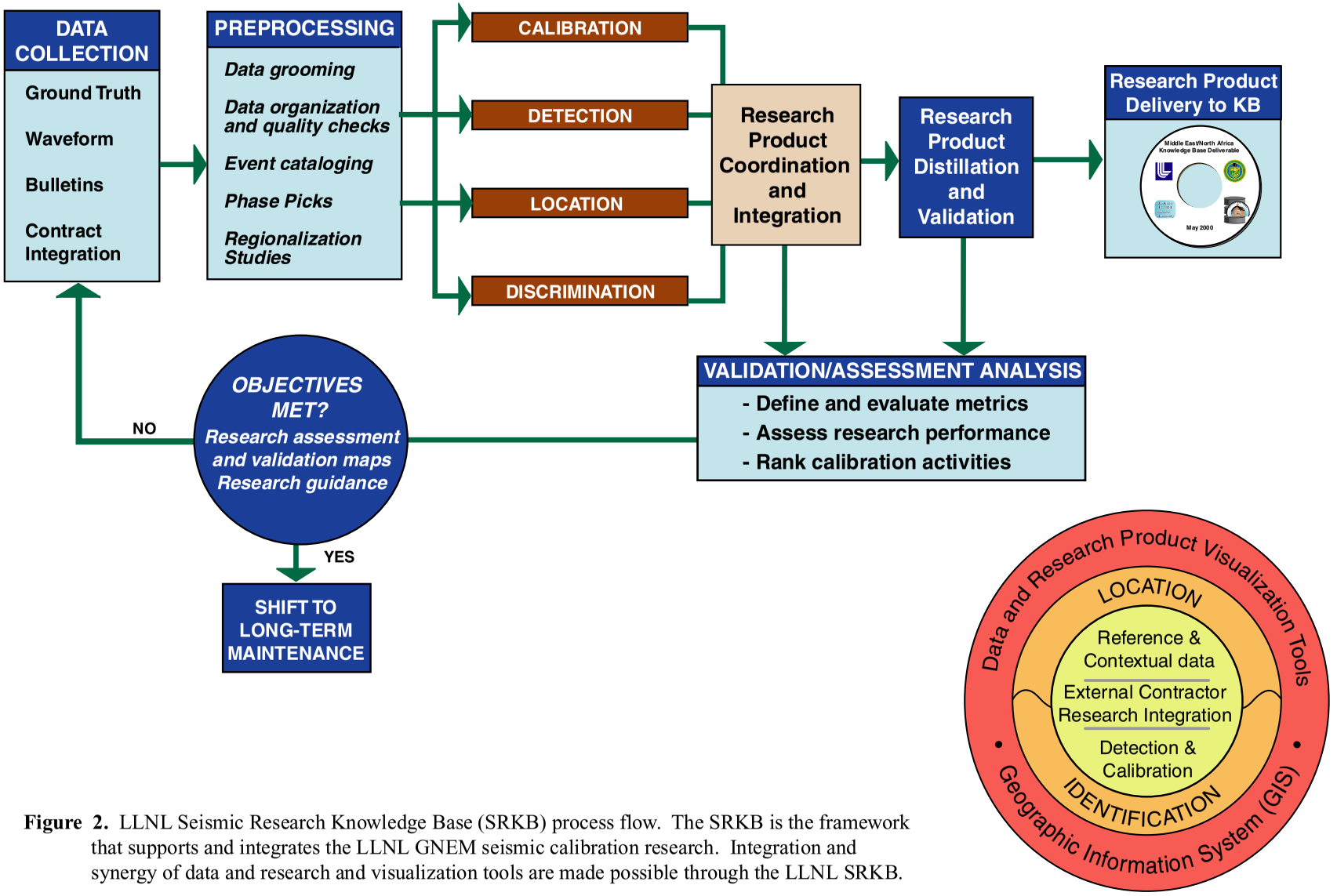

Figure 2. LLNL Seismic Research Knowledge Base (SRKB) process flow. The SRKB is the framework that supports and integrates the LLNL GNEM seismic calibration research. Integration and synergy of data and research and visualization tools are made possible through the LLNL SRKB.

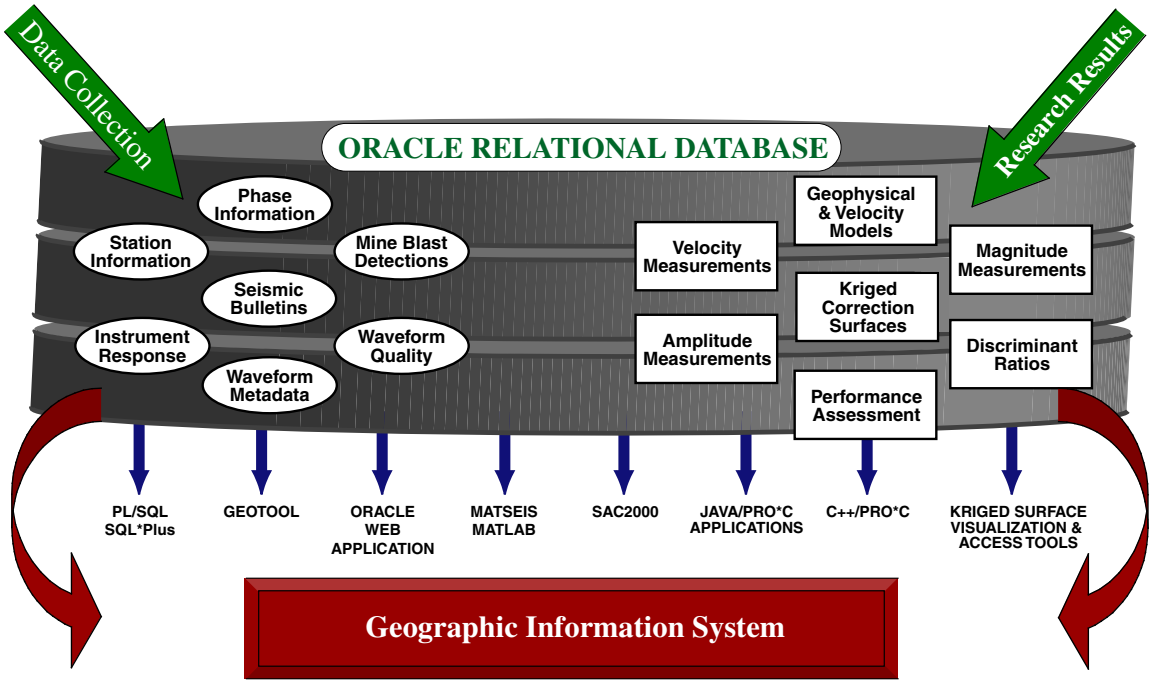

Figure 3. The LLNL SRKB integrates a relational database, a Geographic Information System (GIS) and visualization/data management processes. Many aspects of LLNL GNEM research are contained within the database and can be accessed using a wide variety of tools.

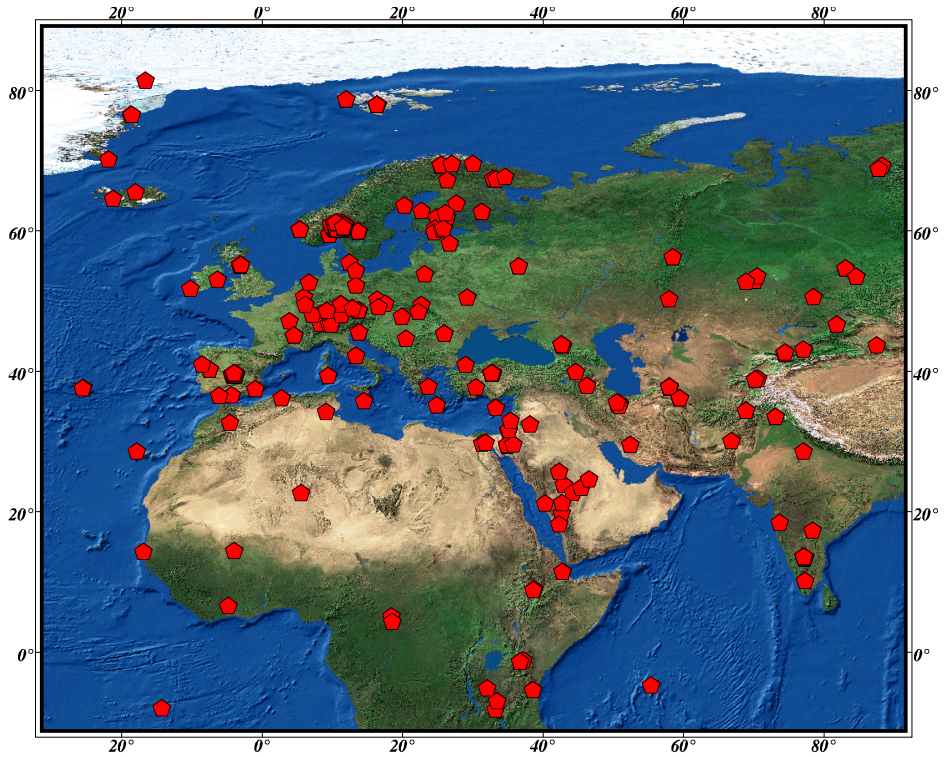

Figure 4. Seismic stations in the ME/NA/WE region with event waveforms stored in the LLNL SRKB.

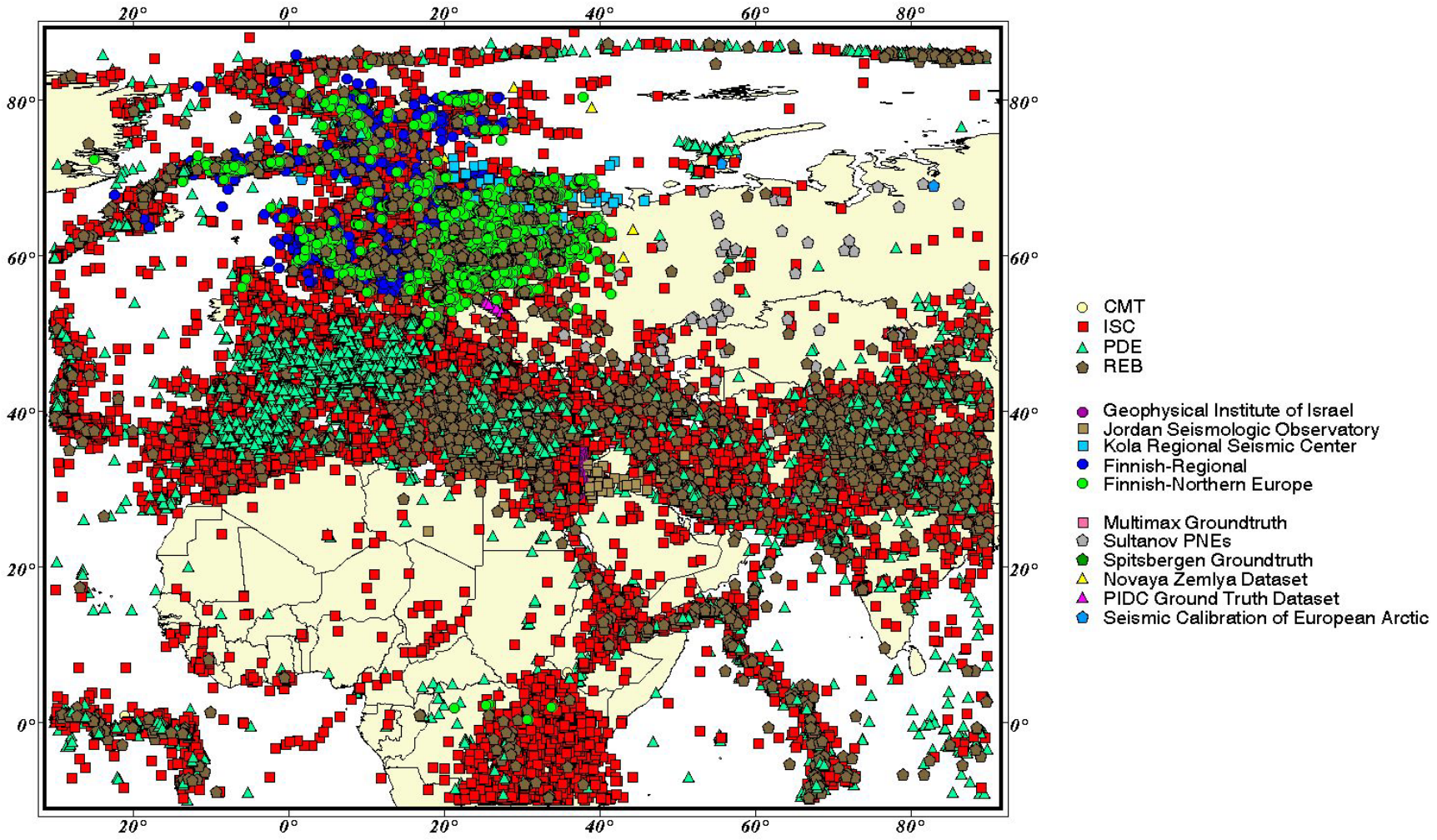

Figure 5. Plot of seismic events in the ME/NA/WE region. The SRKB contains multiple seismic event catalogs that have been reconciled into one database table. These catalogs range in scope from global to regional to special ground truth datasets and provide a much broader range of event magnitudes and event types than any single catalog.